Note: The first five free-response items are worth 4 points each. The sixth item is worth 8 points. Use your discretion in awarding points. The free-response portion of the exam is worth 112 points, so multiply the total points by 4 to calculate the final grade.

1. The midterm grades for two high school algebra classes are provided in the tables below. For convenience, the data have been ordered.

Class 1:

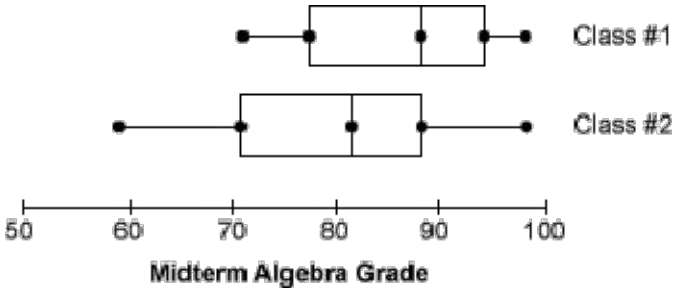
40	**72**	**74**	**75**	**76**	**79**	**82**	**84**	**86**	**91**	**93**
93	**94**	**95**	**95**	**98**	**98**					

Class 2:

59	**65**	**68**	**68**	**69**	**72**	**73**	**78**	**80**	**82**	**82**
83	**83**	**85**	**88**	**88**	**89**	**94**	**96**	**97**	**98**	

A. Sketch parallel boxplots for the two classes.

Answer:



Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E is earned if the graphs are drawn correctly with labels and calculation to check for outliers.

P is earned if the graph is correct but missing labels, or if the graph is correct with labels, but no outlier check is performed.

I for anything else.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

B. Which class has the better overall performance? Justify your answer.

Answer:
Class #1 has the better overall performance. The median or typical grade for Class #1 is 88.5 compared to 82 for Class #2. Both distributions are slightly skewed left, indicating fewer students with low scores; however, the bottom 25% of students in Class #1 had scores from 72 to 77.5 while the bottom 25% from Class #2 had scores from 59 to 70.5. The range of scores for the middle 50% of Class #1 are from 77.5 to 94.5, while the middle 50% of Class #2 scores from 70.5 to 88.5.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the student correctly identifies Class #1 with the better performance and justifies his or her answer by using two out of the following three: center, shape, and spread.

P if the student correctly identifies Class #1 with the better performance, but uses weak justification or only one of the following three: center, shape, and spread.

I if the student correctly identifies Class #1, but without justification, or if he or she chooses Class #2 with the better performance.

C. The teacher for the two classes calculates a 90% confidence interval for the difference in mean midterm grades between Class #1 and Class #2. The confidence interval is (0.09299, 11.413). Assume that confidence interval conditions are met. Based on the confidence interval, do you think there is a difference in the population mean midterm grades for the two classes? Explain.

Answer:
Since zero is not included in the CI (zero indicates there is no difference), there appears to be a difference in the mean midterm grades of the two classes.

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

Scoring for Question 1 (add points from 1A to 1C):			
4 points	**3 points**	**2 points**	**1 point**
EEE	EPE, PEE	EEI, EIE, IEE, EPI, PEI, PPE	EII, IEI, PPI

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

2. A simple random sample of families living in a Chicago suburb was selected. Each family was asked to estimate the number of times each month they went out to eat. The families were told to use a 30-day month and to consider breakfast, lunch, and dinner. Household size and number of times dining out per month are shown in the table below:

Number of Times Dining Out Per Month			
Number in household	**Under 10**	**10–20**	**Over 20**
1 – 3	**12**	**14**	**19**
4 – 6	**25**	**20**	**8**
7+	**46**	**13**	**3**

A. What is the probability that a person chosen at random from the sample will come from a household with 4 – 6 people?

Answer:

P(4 – 6 people in household) $= \frac{53}{160} = 0.33125$

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

B. What is the probability that a person chosen at random from the sample who goes out to eat under 10 times per month will come from a household with 4 – 6 people?

Answer:

Conditional probability: P(4 – 6 people in household who goes out to eat less than 10 times per month) $= \frac{25/160}{83/160} = \frac{25}{83} = 0.3012$

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if calculated correctly with work shown. Arithmetic error is O.K. as long as the work is shown. If the student uses an incorrect answer from part A, but uses correct mechanics, he or she can earn E on this part.

P if the student reverses the condition or uses information from the wrong column.

I for anything else.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

C. Is the number of times per month a family goes out to eat independent of household size? Explain.

Answer:
If the two events are independent, then the answers to parts A and B should be the same, since the condition (going out to eat less than 10 times per month) will have no effect on the outcome of the event within a household of size 4 – 6. Since the probabilities are not the same, the events are not independent.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the student states the events are not independent and gives a correct justification.

P if the student notes that the two events are not independent, but gives no explanation or a weak justification.

I for anything else.

Scoring for Question 2 (add points from 2A to 2C):			
4 points	**3 points**	**2 points**	**1 point**
EEE	EEP, EPP, EPE	IEE, EEI, EIE, EPI, EIP, IEP, IPE	EII, IEI, IIE, IPP

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

3. The Madison Cosmetics Company has developed a new women's facial cleanser that they believe is superior to those currently on the market in the $5 – $10 range. They wish to test the cleanser against the two current best sellers.

A. Identify a potential problem with using a completely randomized experimental design to test the cleanser. Explain.

Answer:
Using a completely randomized design would randomly assign any woman to the different treatment groups, which would prevent the company from seeing if the new cleanser was more effective on a certain age group or a certain skin type, etc. These effects would be impossible to differentiate if all women were randomly assigned to the different treatment groups.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the student identifies the problem and states why it is a problem.

P if the student demonstrates that he or she knows what a completely randomized design is, but doesn't really pinpoint a problem with reason.

I for anything else.

B. The company executives decide to design and carry out a block design experiment to test the new cleanser against its competition. Choose an appropriate blocking scheme for the experiment and justify your choice.

Answer:
Since the purpose of blocking is to form homogenous groups, blocking by either age or skin type would be appropriate since either is likely to have a strong relationship with general skin condition, and the effectiveness of the cleanser could more easily be measured.

Note: Any other reasonable blocking scheme with justification is acceptable for full credit.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the student gives a valid blocking scheme with a plausible reason.

P if the student states a valid blocking scheme, but gives unclear justification or no justification.

I if the student does not give a valid blocking scheme.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

C. Should the company put a control group into the experiment? Why or why not?

Answer:
The company *could* put in a control group, but it is not necessary since there are already two other treatments (the two most popular facial cleansers) in the experiment for purposes of comparison.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the student answers *no* and gives a reason, or if the student answers *yes* but states that there is not really an advantage to doing so.

P if the student demonstrates knowledge of what a control group is and states either *no* or *yes,* but without providing a reason.

I if the student states only *no* or *yes* without providing a reason.

D. Could the experiment be a double-blind experiment? Explain.

Answer:
Yes. A double-blind experiment is one in which neither the individual receiving the treatment nor the individual administering it knows which treatment is being given.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the student states *yes* and shows a clear understanding of what a double-blind experiment is.

P if the student states *yes*, but gives an unclear definition of a double-blind experiment.

I if the student states *no*.

Scoring for Question 3 (add points from 3A to 3D):			
4 points	**3 points**	**2 points**	**1 point**
EEEE, EEEP	EEEI, EEPP	EEPI, EEII, PPPI	EPII, PPII

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

4. Double Dice is a game in which two dice are rolled. A player wins the game if a certain sum is rolled. A player wins $7.00 if he rolls a sum of 5 or 6 and $20.00 if he rolls a sum of 2 or 3. He wins nothing if any other sum is rolled.

A. Define the random variable, *X*.

Answer:
The random variable *X* describes the amount of money won or payout for the game.

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

B. Write out the probability distribution function for the game.

Answer:

x	0	7	20
p(*x*)	2/3	1/4	1/12
or	24/36	9/36	3/36

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

C. What is the expected value of the game?

Answer:

$$E(X) = 0\left(\frac{2}{3}\right) + 7\left(\frac{1}{4}\right) + 20\left(\frac{1}{12}\right) = 0 + 1.75 + 1.67 = \$3.42$$

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

D. Assuming you have to pay $6.00 to play the game, explain what happens in the long run. (Is it a good idea to play the game?)

Answer:
In the long run, the house will keep $2.58 ($6.00 – $3.42) of each $6.00 bet.

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

Scoring for Question 4 (add points from 4A to 4D):			
4 points	**3 points**	**2 points**	**1 point**
EEEE	EEE	EE	E

5. A team of researchers wants to determine whether pet owners are generally more satisfied with their lives than non-pet owners. To test their theory, the researchers randomly select 500 pet owners and 500 non-pet owners from several major metropolitan areas in the country. The researchers then interview the individuals, asking them a series of questions. Each response is assessed with a point value that is later translated to a satisfaction indicator. Of the pet owners surveyed, 380 of the 500 were found to be satisfied with their lives, while 336 of the 500 non-pet owners were found to be satisfied.

A. Would this study be considered an experiment or an observational study? Explain.

Answer:
This is an observational study since no treatment is imposed.

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

E if identified as an observational study with explanation why.

I if identified as an observational study with no reason why, or if identified as an experiment.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

B. Say that the researchers decide to run a *z*-test for the difference of two proportions to see if a higher proportion of pet owners are satisfied with their lives than non-pet owners. What would be the null and alternative hypotheses tested? Be sure to identify the parameters being tested.

Answer:
Note: There are three possible pairs of hypotheses. Nonstandard notation in the hypotheses is acceptable as long as it is defined as a population parameter and not a sample statistic. If nonstandard notation is used and it is not indicated that the notation represents a population parameter, then this part is scored as Incorrect.

Let $p_p =$ the proportion of pet owners who are satisfied with their lives.

Let $p_n =$ the proportion of non-pet owners who are satisfied with their lives.

Null hypothesis: $p_p = p_n$ or $p_p = p_n = 0$ or $p_n - p_p = 0$
Alternative hypothesis: $p_p > p_n$ or $p_p - p_n > 0$ or $p_n - p_p < 0$

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if a correct pair of hypotheses is given and parameters of interest are identified. Note: Any nonstandard notation used must be stated to be a population proportion (otherwise score as I).

P if a correct pair of hypotheses is given (see note about nonstandard notation above) without identifying parameters of interest. Also score a P if incorrect hypotheses are given, but parameters of interest are correctly identified.

I if neither part is done correctly.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

C. What conditions must be checked in order to perform this test? Comment on whether or not each condition is satisfied.

Answer:

1. Independent samples: It is stated in the problem that the samples are random, so this is reasonable.

2. Large sample sizes are verified by checking the following:
   Let $\hat{p}_c = \frac{380+336}{1000} = 0.716$ (combined successes for the two samples)
   Check: $n_p\hat{p}_c \geq 5$ and $n_p(1-\hat{p}_c) \geq 5$ also $n_n\hat{p}_c \geq 5$ and $n_n(1-\hat{p}_c) \geq 5$
   500(0.716) = 358 and 500 (0.284) = 142 same
   So this condition is also met.
   Note: The student may use 10 instead of 5.

   Scoring notes:
   This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

   E if both conditions are mentioned and verified as satisfied. Calculations must be shown for full credit.

   P if the student lists both conditions, but doesn't state they are satisfied or doesn't show calculations.

   I if incorrect conditions are stated.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

D. The *p*-value for the test against the alternative hypothesis, which states that a higher proportion of pet owners are satisfied with their lives than non-pet owners, is 0.0010. Interpret this *p*-value.

Answer:
Assuming the null hypothesis to be true, meaning that pet owners and non-pet owners are equally satisfied with their lives, the *p*-value measures the chance (in this case 0.1%) of observing a difference in the two sample proportions as large or larger than the one observed.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E if the *p*-value is interpreted correctly.

P if the *p*-value interpretation is correct, but doesn't contain "as large or larger" or a similar verbiage. *P* can also be earned if the student makes a decision regarding the null hypothesis using the *p*-value and a significance level, and then gives a conclusion in the context of the problem.

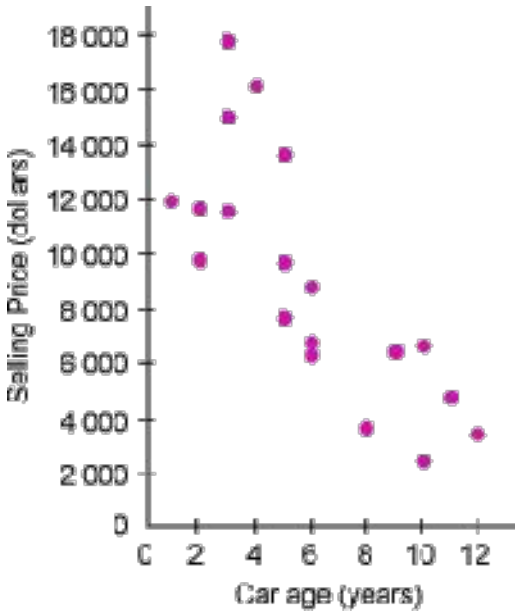
I if *p*-value is incorrectly interpreted or if the student makes a decision regarding the null hypothesis and a significance level with no connection to the context of the problem.

Scoring for Question 5 (add points from 5A to 5D):			
4 points	**3 points**	**2 points**	**1 point**
EEEE	IEEE, EEEP, EEPE, EPEE, EEPP, EPEP, EPPE	IPPP or any combination of 2 E's and 1 P	IPPI, IPIP, IIPP, or any combination of 1 E and 1 P

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

6. The scatterplot below produced the following regression analysis results:

$\bar{x} = 5.8421$ $s_x = 3.3043$ $\bar{y} = 9291.8421$ $s_y = 4387.9983$ $r = -0.7850$



Car age	Price
5	7,888
3	11,650
2	11,774
2	9,900
11	4,995
10	6,800
1	11,988
5	9,900
6	8,950
12	3,500
5	13,700
6	6,850
9	6,750
10	2,500
3	17,900
6	6,700
8	3,800
4	16,000
3	15,000

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

A. Find the equation of the least squares regression line. Show all calculations, and be sure to define any variables used.

Answer:

Slope: $b = r\left(\frac{s_y}{s_x}\right) = -0.7850\left(\frac{4387.9983}{3.3043}\right) = -1042.4534$

y-intercept: $a = \bar{y} - b\bar{x} = 9291.8421 + 1042.4534(5.8421) = 15{,}381.9591$
Equation: $\hat{y} = 15{,}381.9591 - 1042.4534x$

Where *x* = age of car, in years and *y* = selling price

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E for the correct equation of the LSRL with work shown and variables defined.

P for correct equation of LSRL with work shown and variables not defined.

I if the equation is incorrect.

Note: Arithmetic errors that result in an incorrect LSRL should not be counted against the student as long as the correct formulas have been used.

B. Interpret the slope of the regression line in the context of the problem.

Answer:
For each increase of 1 year in a car's age, there is a corresponding decrease of an average of $1,042 in selling price.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E for correct interpretation of slope including use of "on average" or similar expression.

P for correct interpretation without "on average" or similar expression.

I for incorrect interpretation or reversing *x* and *y*.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

C. Calculate and interpret r^2 in the context of the problem.

Answer:
Calculation: $r^2 = (-0.7850)^2 = 0.6162$

Interpretation: About 61.6% of the variation in the selling price of used cars can be accounted for by the linear relationship with age.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E for correct calculation with correct interpretation.

P for correct calculation with unclear or incorrect interpretation.

I for incorrect calculation and incorrect interpretation.

D. Three additional cars are added to the scatterplot: an 11-year-old car that sold for $9,000, a 12-year-old car that sold for $12,000, and a 13-year-old car that sold for $10,000. What effect would these points have on the regression line and correlation? Explain.

Answer:
The student will need to recognize these points as influential, which will result in a change in the position of the regression line. Since the points are influential, they have a tendency to pull the regression line toward them, resulting in the slope of the regression line becoming "less steep" or "more flat." The *y*-intercept will also decrease. The correlation becomes less strong as well.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

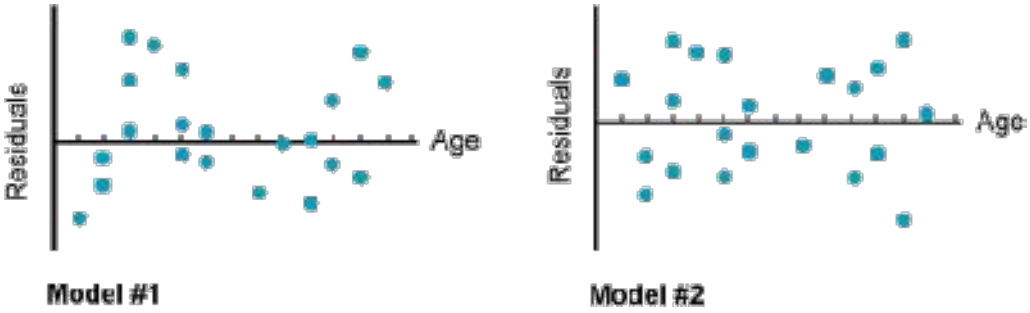
E if slope, *y*-intercept and correlation are all correctly addressed.

P if slope with *y*-intercept or correlation is correctly addressed. Note: If the student states that slope increases without referring to a change in the position of the line and/or states that the correlation increases without specifically referring to strength, grade as P.

I if only slope or *y*-intercept is addressed or if an incorrect interpretation is given.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

E. After adding the vehicles from part D, it is thought that perhaps a nonlinear model would better fit the data. Use the two residual plots shown below to choose a new model to represent the data. Justify your choice.



Model #1 **Model #2**

Answer:
Based on the residual plots shown, the student should choose Model #2 because the residual plot does not appear to show a pattern in contrast to that of Model #1, which appears to have a definite curved pattern.

Scoring notes:
This part is graded as E = essentially correct or I = incorrect.

E for a choice with a correct justification.

I for a choice with incorrect or no justification.

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.

F. Consider cars that are 4 years of age or less. Would the original linear model you found in part A still be appropriate for the data? Explain.

Answer:
With regard to the original linear model, the student should realize that cars 4 years of age or less actually have a positive relationship with their selling price according to the scatterplot. Therefore, the student should come to the conclusion that the linear model in part A would not be appropriate for cars in that age range because of that positive relationship.

Scoring notes:
This part is graded as E = essentially correct, P = partially correct, or I = incorrect.

E for realizing that the linear model would no longer be appropriate and with correct reason.

P for realizing that the linear model would no longer be appropriate with incorrect or no justification.

I if the student states that the linear model is still appropriate.

Scoring for Question 6 (add points from 6A to 6F):	
4 points	**3 points**
E's on parts (A) and (B) E's on parts (C) and (D) E on part (E) E or P on part (F)	E's on parts (A) and (B) E's on parts (C) and (D) E on part (E) or part (F)
2 points	**1 point**
E's on parts (A) and (B), E on part (C) or (D), E on part (E), E or P on part (F) **OR** E on part (A) or (B), E's on part (C) and (D), E on part (E), E or P on part (F) **OR** E on part (A) or (B), E on part (C) or (D), E's on parts (E) and (F)	2 E's and 1 or 2 P's **OR** 1 E and 4 or 5 P's **OR** 5 P's

© Copyright 2009 Apex Learning Inc. All rights reserved. This material is intended for the exclusive use of registered users only. No portion of these materials may be reproduced or redistributed in any form without the express written permission of Apex Learning Inc.